The Ewens process on spaces of even and balanced partitions

Harry Crane and Peter McCullagh

October 19, 2011

[1]

Abstract

We discuss a generalization of the Ewens partition process to $\mathcal{P}_{\mathbb{N}:j}$, the space of partitions whose block sizes are divisible by $j \in \mathbb{N}$, called even partitions of order j, or j-even partitions, and $\mathcal{P}^0_{\mathbb{N}:j}$, the subspace of $\mathcal{P}_{\mathbb{N}:j}$ whose elements are labeled in $[j]$ and whose blocks contain an equal number of elements with each label, called j-balanced partitions. As in the Ewens process, these processes can be constructed sequentially according to a random seating rule. For both balanced and even partitions we can associate a projective system on which we construct random processes which project to the Chinese restaurant processes on the respective partition spaces.

1 Preliminaries

In this paper, we discuss probability distributions on projective systems of set partitions and permutations. We now introduce notation and terminology which we need in our development.

A permutation σ of a set $A \subset \mathbb{N}$ is a one-to-one and onto map $A \to A$. For $n \geq 1$, $\mathcal{S}_n$ denotes the symmetric group of permutations $[n] \to [n]$ and $\mathcal{S}$ denotes the space of permutations of $\mathbb{N}$. For any permutation σ, we write $\#\sigma$ to denote the number of cycles of σ.

To each permutation σ of $A \subset \mathbb{N}$ the mapping $\pi : \mathcal{S}_n \to \mathcal{P}_{[n]}$ takes $\sigma \mapsto \pi(\sigma)$, the partition of A whose blocks are given by the cycles of σ. Throughout this paper, $\mathcal{P}$ denotes the space of set partitions of the natural numbers $\mathbb{N}$. We regard an element B of $\mathcal{P}$ as a collection of disjoint non-empty subsets, called blocks, written $B = \{B_1, B_2, \ldots\}$, such that $\bigcup_i B_i = \mathbb{N}$. For $B \in \mathcal{P}$ and $b \in B$, $\#B$ is the number of blocks of B and $\#b$ is the number of elements of b. Wherever necessary, $\mathcal{P}^{(k)}$ denotes the space of partitions of $\mathbb{N}$ with at most k blocks, i.e. $\mathcal{P}^{(k)} := \{B \in \mathcal{P} : \#B \leq k\}$. For any $A \subset \mathbb{N}$, let $B_{|A}$ denote the restriction of B to A, obtained by deleting the complement of A from the index set, i.e. for $B = \{B_1, B_2, \ldots\}$ and $A \subset \mathbb{N}$, $B_{|A} := \{B_i \cap A : i \geq 1\}$. For fixed $n \in \mathbb{N}$, $\mathcal{P}_{[n]}$ and $\mathcal{P}^{(k)}_{[n]}$ are the restriction to $[n] := \{1, \ldots, n\}$ of $\mathcal{P}$ and $\mathcal{P}^{(k)}$ respectively.

For each $n \geq 1$, a partition B of $[n]$ determines an integer partition of n by its block sizes. An integer partition $\lambda = (\lambda_1, \ldots, \lambda_n)$ of $n \in \mathbb{N}$ is a list of multiplicities, where λ_j is the number of parts

[1]Research supported by NSF grant no. DMS-0305009

of size j, sometimes also written as $1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n}$, such that $\sum_{j=1}^n j\lambda_j = n$. The number of parts of an integer partition is given by the sum $\lambda_\bullet := \sum_{j=1}^n \lambda_j$. We write $\mathcal{P}_n$ to denote the space of integer partitions of n.

For a partition $B \in \mathcal{P}_{[n]}$, $\nu : \mathcal{P}_{[n]} \to \mathcal{P}_n$ maps B to the integer partition induced by its block sizes, $B \mapsto (\lambda_1, \ldots, \lambda_n)$ where $\lambda_j := \#\{b \in B : \#b = j\}$. Hence, we have the following relationship among the spaces $\mathcal{S}_n$, $\mathcal{P}_{[n]}$ and $\mathcal{P}_n$.

$$\mathcal{S}_n \to_\pi \mathcal{P}_{[n]} \to_\nu \mathcal{P}_n.$$

To each integer partition $\lambda \in \mathcal{P}_n$ there are $\frac{n!}{\prod_{j=1}^n (j!)^{\lambda_j}\lambda_j!}$ set partitions of $[n]$ whose block sizes correspond to the parts of λ and to each set partition $\pi \in \mathcal{P}_{[n]}$ there are $\prod_{b\in\pi}\Gamma(\#b)$ permutations of $[n]$ whose cycles correspond to the blocks of π. Hence, given a probability distribution on integer partitions of n, we easily obtain a probability distribution on $\mathcal{P}_{[n]}$ and $\mathcal{S}_n$ by first sampling a random integer partition λ and, conditional on λ, choosing uniformly among the elements of either $\mathcal{P}_{[n]}$ or $\mathcal{S}_n$ which correspond to λ. The probability of a partition or permutation obtained in this way is obtained by multiplying by the appropriate combinatorial factor.

Ewens [5] first introduced his sampling formula as a distribution on $\mathcal{P}_n$. For $\alpha > 0$, the Ewens(α) distribution on $\mathcal{P}_n$ is

$$p_n(\lambda; \alpha) = \frac{n!}{\alpha^{\uparrow n}} \prod_{j=1}^n \frac{\alpha^{\lambda_j}}{j^{\lambda_j}\lambda_j!}, \tag{1}$$

where $\alpha^{\uparrow n} := \alpha(\alpha+1)\cdots(\alpha+n-1)$. However, the collection $(\mathcal{P}_n, n \geq 1)$ is not a projective system and so these distributions do not determine a process on an infinite space. Our treatment relies on specifying a family of probability distributions on a projective system, which we now define.

A projective system associates with each finite set $[n]$ a set Q_n and with each one-to-one injective map $\varphi : [m] \to [n]$, $m \leq n$, a projection $\varphi^* : Q_n \to Q_m$ which maps Q_n into Q_m such that

- if φ is the identity $[n] \to [n]$ then φ^* is the identity $Q_n \to Q_n$ and
- if $\psi : [l] \to [m]$, $l \leq m$, $\psi^* : Q_m \to Q_l$ is its associated projection, the composition $(\varphi\psi) : [l] \to [n]$ satisfies $(\varphi\psi)^* \equiv \psi^*\varphi^* : Q_n \to Q_l$.

If Q_n is the set of subsets of $[n]^2$, i.e. the space of directed graphs with n vertices, one can define the projection $Q_n \to Q_m$ either by *restriction* or *delete-and-repair*. Each $A \in Q_n$ can be represented as an $n \times n$ matrix with entries in $\{0, 1\}$ such that $A_{ij} = 1$ if $(i, j) \in A$ and $A_{ij} = 0$ otherwise. For each $n \geq 1$, let $D_{n,n+1}$ be the operation on Q_{n+1} which restricts A to the complement of $\{n + 1\}$. In matrix form, $D_{n,n+1}A =: A_{|[n]}$ is the $n \times n$ matrix obtained from A by removing the last row and last column of A and keeping the rest of the entries unchanged. It is clear that the compositions $D_{m,n} := D_{m,m+1} \circ \cdots \circ D_{n-1,n}$ for $m \leq n$ are well-defined as the restriction of $A \in Q_n$ to $[m]$ by removing the last $n - m$ rows and columns of A. The restriction maps $(D_{m,n}, m \leq n)$ together with permutation maps $(\sigma \in \mathcal{S}_n, n \geq 1)$ make $(Q_n, n \geq 1)$ a projective system.

Another way to specify a projective system on $(Q_n, n \geq 1)$ is by *delete-and-repair*. For $n \geq m \geq 1$, let ψ_m act on $A \in Q_n$ by removing the mth row and column of A and directing an edge from each i in $\{j \in [n] : (j, m) \in A\}$ to each k in $\{j \in [n] : (m, j) \in A\}$. In other words, $\psi_m A$ is obtained by deleting the vertex labeled m from A and connecting two vertices i and k by a directed

edge from i to k if both (i,m) and (m,k) are elements of A, i.e. there is a directed path $i \to m \to k$ in A.

For $m \leq n$, define $\psi_{m,n} := \psi_{m+1} \circ \cdots \circ \psi_n$. Plainly, $\psi_{m,n}$ is well-defined since for each $n \geq 2$, $\psi_{n-2,n} \equiv \psi_{n-1} \circ \psi_n = \psi_n \circ \psi_{n-1}$ and $\psi_{l,n} = \psi_{l,m} \circ \psi_{m,n}$.

The projections $(D_{m,n}, m \leq n)$ and $(\psi_{m,n}, m \leq n)$ defined above define two different projective systems on $(Q_n, n \geq 1)$. Our discussion involves the systems $\left(\mathcal{P}_{[n]}, n \geq 1\right)$ of finite set partitions and $(\mathcal{S}_n, n \geq 1)$ of finite set permutations. Both of these can be thought of in the context of directed graphs. A partition $B \in \mathcal{P}_{[n]}$ can be regarded as a subset of $[n]^2$ such that $(i,j) \in B$ if and only if i and j are in the same block of B. Thus, $\left(\mathcal{P}_{[n]}, n \geq 1\right)$ together with restriction maps $(D_{m,n}, m \leq n)$ and permutation maps is a projective system. In fact, the restriction and delete-and-repair projections are equivalent on $\left(\mathcal{P}_{[n]}, n \geq 1\right)$.

A permutation $\sigma \in \mathcal{S}_n$ is a one-to-one and onto map $[n] \to [n]$ and we regard σ as a subset of $[n]^2$ by $(i,j) \in \sigma$ if $\sigma(i) = j$. For $\sigma \in \mathcal{S}_{n+1}$, delete-and-repair acts on σ by $\sigma' := \psi_{n,n+1}\sigma$ which satisifies

$$\sigma'(i) = \begin{cases} \sigma(n+1), & i = \sigma^{-1}(n+1) \\ \sigma(i), & \text{otherwise.} \end{cases}$$

Hence, $\sigma' \in \mathcal{S}_n$ and $(\mathcal{S}_n, n \geq 1)$ together with delete-and-repair maps $(\psi_{m,n}, m \leq n)$ and permutation maps is a projective system.

A random partition $B \in \mathcal{P}$ is a probability distribution μ on $\mathcal{P}$, also referred to as a partition structure by Kingman [6]. By Kolmogorov's extension theorem [2], any μ on $\mathcal{P}$ is uniquely characterized by its collection of finite-dimensional distributions $(\mu_n, n \geq 1)$ on the projective system $\left(\mathcal{P}_{[n]}, n \geq 1\right)$ if and only if

- μ_n is invariant under action of $\mathcal{S}_n$ on $[n]$ for each $n \geq 1$;
- μ_n is the marginal distribution of μ_{n+1} under $D_{n,n+1}$, i.e. $\mu_n \equiv \mu_{n+1} D^{-1}_{n,n+1}$.

The first condition is called finite exchangeability and the second consistency, or compatibility. Any collection $(\mu_n, n \geq 1)$ of finitely exchangeable and consistent measures on a projective system $(Q_n, n \geq 1)$ defines an infinitely exchangeable measure μ on the infinite space Q which $(Q_n, n \geq 1)$ characterizes.

The conditions of finite exchangeability and compatibility are also natural in the context of statistical models. In statistical applications, individuals in an infinite population are indexed by $\mathbb{N}$ and a finite sample $[n]$ from this population is obtained. Finite exchangeability ensures that any inference drawn by applying the model $(\mu_n, n \geq 1)$ will not depend on the arbitrary labels which are assigned to the sampled units. Compatibility ensures that inference is consistent under subsampling. Though these are natural conditions, constructing a family of distributions on $\left(\mathcal{P}_{[n]}, n \geq 1\right)$ for which they are satisfied is non-trivial. Kingman [6] characterized all infinitely exchangeable random partitions through what is known as the paintbox process. We discuss two particular infinitely exchangeable families throughout this paper, the Ewens family and its close relative, the (α, θ)-model.

Kingman [6] showed that the Ewens distribution sits more naturally on the projective system $(\mathcal{P}_{[n]}, n \geq 1)$ of set partitions with restriction maps $(D_{m,n}, m \leq n)$ and permutation maps. For each

$n \geq 1$, the finite-dimensional Ewens(α) distributions on set partitions are given by

$$P_n(B; \alpha) = \alpha^{\#B}/\alpha^{\uparrow n} \prod_{b \in B} \Gamma(\#b), \tag{2}$$

for $\alpha > 0$. The distribution $P_n(\cdot; \alpha)$ is obtained from (1) by dividing by the factor $\frac{n!}{\prod_{j=1}^{n} (j!)^{\lambda_j} \lambda_j!}$, as discussed above. As these finite-dimensional distributions are finitely exchangeable and consistent on a projective system, Kolmogorov's extension theorem [2] guarantees that they characterize an infinitely exchangeable partition process on $\mathcal{P}$.

Furthermore, dividing by the factor $1/\prod_{b \in B} \Gamma(\#b)$ yields the Ewens(α) distribution on permutations

$$Q_n(\sigma; \alpha) = \alpha^{\uparrow \#\sigma}/\alpha^{\uparrow n},$$

which characterizes an infinitely exchangeable process on the projective system $(\mathcal{S}_n, n \geq 1)$ of permutations with delete-and-repair maps $(\psi_{m,n}, m \leq n)$.

Pitman [8] describes a two-parameter model, called the (α, θ)-model, whereby $B := (B_n, n \in \mathbb{N})$, a sequence of partitions with $B_n \in \mathcal{P}_{[n]}$, is constructed sequentially as follows.

After the first n integers have formed partition B_n of $[n]$, element $n + 1$ is

- inserted in $b \in B_n$ with probability $(\#b - \alpha)/(n + \theta)$;
- appended to B_n as a singleton block with probability $(\theta + \alpha\#B_n)/(n + \theta)$.

B_n obtained is this way is said to follow a Chinese restaurant process CRP(n, α, θ) with parameter (α, θ).

For (α, θ) which satisfies either

- $\alpha = -\kappa < 0$ and $\theta = m\kappa$ for some $m \in \mathbb{N}$ or
- $0 \leq \alpha \leq 1$ and $\theta > -\alpha$

the above seating rule characterizes an infinitely exchangeable process on $\mathcal{P}$ with finite-dimensional distributions

$$P_n(B; \alpha, \theta) = \frac{(\theta/\alpha)^{\uparrow \#B}}{\theta^{\uparrow n}} \prod_{b \in B} -(-\alpha)^{\uparrow \#b}. \tag{3}$$

Note that for $\alpha = -\kappa < 0$ and $\theta = m\kappa$, letting $m \to \infty$ and $\theta \to \lambda > 0$ in (3) gives the Ewens(λ) distribution (2).

In this paper, we discuss an extension of the (α, θ)-model to even and balanced partitions.

Definition 1.1. *For $j, n \in \mathbb{N}$, a partition $\sigma \in \mathcal{P}_{[nj]}$ is said to be* even of order j, *or j-even, if the block size of each $b \in \sigma$ is a multiple of j. Let $\mathcal{P}_{[nj]:j}$ denote the set of partitions of $[nj]$ that are even of order j and $\mathcal{P}^{(k)}_{[nj]:j}$ denote the set partitions of $[nj]$ with at most k blocks that are even of order j.*

Let J be a set of types or marks attached to elements in the index set such that $\#J = j$. A j-even partition $\sigma \in \mathcal{P}_{[nj]}$ is said to be balanced of order j*, or j-balanced, if its blocks, $b_1, \ldots, b_{\#\sigma}$, each contain an equal number of elements of each type. Let $\mathcal{P}^0_{[nj]:j}$ denote the set of j-balanced partitions $[nj]$ and $\mathcal{P}^{0|(k)}_{[nj]:j}$ denote the set of j-balanced partitions of $[nj]$ with at most k blocks.*

For example, for $j = 3$, $135789|246 \in \mathcal{P}_{[9]:3}$ and $1_a 3_c 5_b 7_a 8_b 9_c | 2_b 4_a 6_c \in \mathcal{P}^0_{[9]:3}$, where each element is labeled in $\{a, b, c\}$.

For each $n \in \mathbb{N}$, the sets $(\mathcal{P}_{[nj]:j}, \leq)$ and $(\mathcal{P}^0_{[nj]:j}, \leq)$, where $\leq$ is the partial ordering known as *sub-partition*, are partially ordered sets, but are not lattices as any two elements of these spaces do not have a unique infimum under $\leq$.

For $j = 2, 3, \ldots$, the systems $\left(\mathcal{P}_{[nj]:j}\right)_{n\geq 1}$ and $\left(\mathcal{P}^0_{[nj]:j}\right)_{n\geq 1}$ do not constitute projective systems under restriction. In what follows, we proceed in three steps for both balanced and even partitions separately. We first construct a projective system which we can associate with the space of either balanced or even partitions and we define a family of finite-dimensional distributions on this system. Finally, we compute the induced distribution on balanced and even partitions by projection.

2 Balanced partitions

We now discuss balanced partitions, which can be related to complete and balanced block designs in the applied statistics literature. In a balanced partition, there are j types and each element is assigned exactly one type. In the design context, the elements represent statistical units and the types are treatments. So each unit is assigned a treatment, each treatment occurs in each block, making the design complete, and each treatment appears the same number of times within each block, making it balanced. Note that blocks need not be of the same size. The interested reader is referred to Bailey [1], which gives a detailed account of experimental design.

2.1 Associated integer partitions

Let $j, n \in \mathbb{N}$ and put $[nj]^* = [n] \times \{(1), \ldots, (j)\}$, the index set which corresponds to marked populations $[n]^{(1)}, \ldots, [n]^{(j)}$, each of equal size. A balanced partition of $[nj]$ can be generated in two steps.

(i) Generate $\pi \sim (\alpha, \theta)$ and $\sigma_2, \ldots, \sigma_j$ i.i.d. uniform matchings such that $\sigma_i : [n]^{(1)} \to [n]^{(i)}$ so

$$\mathbb{P}_{\alpha,\theta}(\pi, \sigma) = \frac{(\theta/\alpha)^{\uparrow\#\pi}}{(n!)^{j-1}\theta^{\uparrow n}} \prod_{b\in\pi} -(-\alpha)^{\uparrow\#b}.$$

(ii) Write $\pi := \{\pi_i\}$ and regard σ_1 as the identity map $[n]^{(1)} \to [n]^{(1)}$. Then $\pi' = \{\cup_{k=1}^{j} \sigma_k \pi_i\}$ is a balanced partition of $[nj]$ and has probability

$$\mathbb{P}_{\alpha,\theta}(\pi') = \frac{(\theta/\alpha)^{\uparrow\#\pi'}}{(n!)^{j-1}\theta^{\uparrow n}} \prod_{b\in\pi'} -(-\alpha)^{\uparrow\#b/j} [(\#b/j)!]^{j-1}.$$

The partition π' obtained in this way determines an integer partition $(m_1, \ldots, m_{nj})$ of nj such that $m_i = 0$ for $i \notin \{j, 2j, \ldots, nj\}$. To each such integer partition, there are $\frac{(n!)^j}{\prod_{i=1}^n (i!)^{jm_{ij}} (m_{ij}!)^{j-1}}$ associated balanced partition with block sizes given by (m_i). Hence, we have

$$\mathbb{P}_{\alpha,\theta}(m_1, \ldots, m_{nj}) = \frac{n!(\theta/\alpha)^{\uparrow m_\bullet}}{\theta^{\uparrow n}} \prod_{i=1}^{n} \frac{[-(-\alpha)^{\uparrow i}]^{m_{ij}}}{(i!)^{m_{ij}} (m_{ij}!)^{j-1}} \tag{4}$$

as the distribution on integer partitions associated with balanced partitions of $[nj]$. Note that this corresponds to the distribution of an (α, θ)-partition of $[nj]$ conditioned to be balanced.

2.2 A Chinese restaurant construction for balanced partitions

A balanced partition of $[nj]$ can be obtained via a Chinese restaurant construction similar to that mentioned for the (α, θ)-model in section 1.

Let J be a set of $j \in \mathbb{N}$ types. Each unit is classified as exactly one type. Without loss of generality, assume $J = [j]$ and that individuals arrive in such a way that the $(mj + i)$th labeled unit is of type i for all $m \in \mathbb{N} \cup \{0\}$. A collection of units $\{u_{i_1}, \ldots, u_{i_{mj}}\}$ is said to be balanced if it contains an equal number of units with each label $j \in J$. We construct $\sigma \in \mathcal{P}^0_{[nj]:j}$ according to the following random seating rule. Let (α, θ) lie in the parameter space of Pitman's two-parameter model, (3).

(1) Units arrive in balanced groups of size j;

(2) the first j units, $u_1, \ldots, u_j$, are seated at the same table;

(3) after nj arrivals are seated according to partition π, the next j units, $u_{nj+1}, \ldots, u_{(n+1)j}$ are seated as follows:

   a. each unit u_{nj+i} for $j \geq 2$, chooses a unit $u^{(i)}$ uniformly among $u_i, u_{j+i}, \ldots, u_{(n-1)j+i}, u_{nj+i}$ and switches positions with $u^{(i)}$ (note: if $u^{(i)} = u_{nj+i}$ there is no change);
   b. $(u_{nj+1}, u^{(2)}, \ldots, u^{(j)})$ is treated as a single unit and chooses a table, $b \in \pi$, according to the two-parameter CRP with parameter (nj, α, θ);
   c. $(u_{nj+1}, u^{(2)}, \ldots, u^{(j)})$ is seated at b.

For $n \geq 1$ and $\pi \in \mathcal{P}^0_{[nj]:j}$ such that $u_1, \ldots, u_{nj}$ are configured according to π at the beginning of step 3, any $\pi' \in \mathcal{P}^0_{[(n+1)j]:j}$ for which there is a realization of steps 3a.-3c. such that $B_{n+1} = \pi'$ with positive probability is said to be accessible from π, written $\pi \mapsto \pi'$.

Proposition 2.1. *The probability distribution on $\mathcal{P}^0_{[nj]:j}$ which results from the above construction is*

$$p^j_n(\pi; \alpha, \theta) = \frac{(\theta/\alpha)^{\uparrow \#\pi}}{(\theta/j)^{\uparrow n} (n!)^{j-1}} \prod_{b \in \pi} -(-\alpha/j)^{\uparrow(\#b/j)} [(\#b/j)!]^{j-1} \tag{5}$$

Proof. We show (5) by induction. Fix $j \in \mathbb{N}$ and (α, θ) subject to the constraints of the two-parameter model. Then (5) holds for $n = 1$ since $p^j_1(\pi; \alpha, \theta) = \mathbb{I}_{\{\#\pi=1\}}$, as it must according to step (1) of the above seating plan.

Now, assume that (5) holds for $n \in \mathbb{N}$ and consider $\pi' \in \mathcal{P}^0_{[(n+1)j]:j}$. Let $A_{\pi'} := \{\pi \in \mathcal{P}^0_{[nj]:j} : \pi \mapsto \pi'\}$. By construction, the block sizes of each $\pi \in A_{\pi'}$ and π' are identical except for the block to which the new group, $(u_{nj+1}, u^{(2)}, \ldots, u^{(j)})$, after appropriate reseating, is inserted in π to obtain π'. Let this block be written as $b^* \in \pi$.

There are $\left(\frac{\#b^*+j}{j}\right)^{j-1}$ partitions in $A_{\pi'}$. To see this, note that for each $i = 2, \ldots, j$, there are $\#b^*/j + 1$ elements of b^* with label i with which u_{nj+i} can choose to switch positions. Each one of these choices has probability $1/(n+1)$.

In step 3a. the group $(u_{nj+i}, u^{(2)}, \ldots, u^{(j)})$ chooses $b \in \pi$ with probability $\frac{\#b-\alpha}{\theta+nj}$ and is placed at a new block with probability $\frac{\theta+\alpha\#\pi}{\theta+nj}$. $\square$

For $\alpha = -\kappa < 0$ and $\theta = m\kappa$, $p^j_n(\cdot;\alpha,\theta)$ in (5) has limit

$$p^j_n(\pi;\lambda) = (\lambda/j)^{\#\pi} \frac{\prod_{b\in\pi} (\#b/j)^{j-1}\Gamma(\#b/j)^j}{(\lambda/j)^{\uparrow n}(n!)^{j-1}},, \tag{6}$$

as $m \to \infty$ and $\theta \to \lambda$, the distribution on $\mathcal{P}^0_{[nj]:j}$ obtained if seating is done according to CRP(nj, λ) in step 3b.

Note that the finite dimensional distribution in (5) is the distribution of the neutral partition determined by the random seating rule above and forgetting the group, $(u_{nj+1}, u^{(2)}, \ldots, u^{(j)})$, to which each unit is associated. This distribution also coincides with the distribution of an (α, θ)-partition conditioned to be balanced, shown in section 2.1. For each $n \geq 1$, these distributions are defined on $\mathcal{P}^0_{[nj]:j}$, which is not a projective system. We now show a projective system associated with balanced partitions on which we define a consistent family of probability distributions.

2.3 A projective system associated with balanced partitions

Let $j, n \in \mathbb{N}$ and consider the two-step construction at the beginning of section 2.1. The distribution of (π, σ) obtained from this procedure is

$$p_n(\pi, \sigma; \alpha, \theta) = \frac{(\theta/\alpha)^{\#\pi}}{(n!)^{j-1}\theta^{\uparrow n}} \prod_{b\in\pi} -(-\alpha)^{\uparrow\#b}. \tag{7}$$

We have shown that the balanced partition determined by π and $(i, \sigma_2(i), \ldots, \sigma_j(i))$ is distributed as (5), but this family of distributions is not defined on a projective system. Each j-tuple $(\pi, \sigma_2, \ldots, \sigma_j)$ obtained from the two-step procedure of section 2.1 is an element in the product space $\mathcal{P}_{[n]} \times \prod_{i=2}^j \mathcal{S}_n$, which is a product of projective systems from section 1. Hence, we define projection maps $\tilde{D}_{m,n} : \mathcal{P}_{[n]} \times \prod_{i=2}^j \mathcal{S}_n \to \mathcal{P}_{[m]} \times \prod_{i=2}^j \mathcal{S}_m$ by $\tilde{D}_{m,n} := (D_{m,n}, \psi_{m,n}, \ldots, \psi_{m,n})$ where $D_{m,n}$ and $\psi_{m,n}$ are, respectively, the *restriction* and *delete-and-repair* maps defined in section 1. The maps $(\tilde{D}_{m,n}, m \leq n)$ act componentwise, i.e. $\tilde{D}_{l,n} = \tilde{D}_{l,m} \circ \tilde{D}_{m,n} \equiv (D_{l,m} \circ D_{m,n}, \psi_{l,m} \circ \psi_{m,n}, \ldots, \psi_{l,m} \circ \psi_{m,n})$ for all $l \leq m \leq n$ and make $\left(\mathcal{P}_{[n]} \times \prod_{i=2}^j \mathcal{S}_n, n \geq 1\right)$ into a projective system.

It is clear that the finite-dimensional distributions on $\mathcal{P}_{[n]} \times \prod_{i=2}^j \mathcal{S}_n$ in (7) are invariant under

the natural action of each $(\sigma_1, \ldots, \sigma_k) \in \prod_{i=1}^{j} \mathcal{S}_n$ on $\mathcal{P}_{[n]} \times \prod_{i=2}^{j} \mathcal{S}_n$. The following establishes consistency with respect to $\tilde{D}_{m,n}$.

Proposition 2.2. *The family of finite-dimensional distributions in* (7) *is consistent with respect to* $\tilde{D}_{m,n}$ *defined above.*

Proof. The distribution in (7) is a product of the (α, θ)-distribution and $j-1$ uniform random permutations of $[n]$. Hence, $(\sigma, \pi) \sim p_n^j(\cdot; \alpha, \theta)$ is a collection of independent random objects, each of which is consistent, so $p_n^j(\cdot; \alpha, \theta)$ is consistent. □

The following is then an immediate consequence of Kolmogorov's theorem [2].

Theorem 2.3. *There exists an infinitely exchangeable measure* p^j *on* $\mathcal{P} \times \prod_{i=2}^{j} \mathcal{S}$ *such that for* $p_n^j(\cdot; \alpha, \theta)$ *in* (7)

$$p_n^j((\pi, \sigma); \alpha, \theta) = p^j(\{(\pi^\infty, \sigma^\infty) : (\pi^\infty, \sigma^\infty)_{|[n]} = (\pi, \sigma)\})$$

for all $(\pi, \sigma) \in \mathcal{P}_{[n]} \times \prod_{i=2}^{j} \mathcal{S}_n$.

Note that we use the notation $(\pi^\infty, \sigma^\infty)_{|[n]}$ to denote the obvious restriction of each component to $[n]$.

Theorem 2.3 establishes the existence of an infinitely exchangeable measure on $\mathcal{P} \times \prod_{i=2}^{j} \mathcal{S}$. For a sequence $(\sigma_n, \pi_n)_{n \geq 1}$ of finite-dimensional restrictions distributed as (7), the sequence $(\sigma_n \pi_n, n \geq 1)$ which ignores the matchings has finite-dimensional distributions as in (5).

2.4 Balanced permutations

The correspondence between partitions of $[n]$ and permutations of $[n]$ discussed in section 1 induces the notion of a balanced permutation. To each partition $\pi \in \mathcal{P}_{[n]}$, there are $\prod_{b \in \pi} \Gamma(\#b)$ associated permutations of $[n]$. Hence, we obtain a distribution on $\prod_{i=1}^{n} \mathcal{S}_n$ from (7) by choosing uniformly among the set of permutations which correspond to $\pi \in \mathcal{P}_{[n]}$. Writing σ_1 for this permutation and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_j)$, we obtain

$$p_n(\sigma; \alpha, \theta) = \frac{(\theta/\alpha)^{\uparrow \#\sigma_1}}{(n!)^{j-1} \theta^{\uparrow n}} \prod_{b \in \sigma_1} \frac{-(-\alpha)^{\uparrow \#b}}{\Gamma(\#b)}.$$

In this specification, we assume that the elements of $(i, \sigma_2(i), \ldots, \sigma_j(i))$ are arranged in order $i, \sigma_2(i), \ldots, \sigma_j(i)$. In this way, the permutation obtained from $(\sigma_1, \ldots, \sigma_j)$ is balanced in the sense that each cycle contains each type an equal number of times and within each block the elements labeled k, for each k, are at distance j from the next closest element labeled k.

We can further randomize the arrangement of the elements by letting $\sigma_0 \in \mathcal{S}_j$ be a uniform permutation of $[j]$ such that $\sigma_0 = 1$, i.e. σ_0 is a cyclic permutation. The permutation σ_0 then specifies the arrangement of elements within groups and the distribution of $(\sigma_0, \sigma_1, \ldots, \sigma_j)$ is

$$\frac{(\theta/\alpha)^{\uparrow \#\sigma_1}}{(j-1)!(n!)^{j-1} \theta^{\uparrow n}} \prod_{b \in \sigma_1} \frac{-(-\alpha)^{\uparrow \#b}}{\Gamma(\#b)}.$$

As in section 2.3, $\prod_{i=1}^{j} \mathcal{S}_n$ and $\mathcal{S}_j \times \prod_{i=1}^{j} \mathcal{S}_n$ are projective systems under $\tilde{\psi}_{m,n} := (\psi_{m,n}, \ldots, \psi_{m,n})$ and $\tilde{\psi}^*_{m,n} := (\mathrm{id}, \tilde{\psi}_{m,n})$ respectively.

3 Even partitions

Let $j, n \in \mathbb{N}$, we generate an even partition of $[jn]$ in two steps just as in section 2.

(i) Generate $\pi \sim (\alpha, \theta)$ and $\sigma \in \mathcal{S}_{nj}$, a uniform permutation of $[nj]$ so that

$$\mathbb{P}_{\alpha,\theta}(\pi, \sigma) = \frac{(\theta/\alpha)^{\#\pi}}{(nj)!\theta^{\uparrow n}} \prod_{b\in\pi} -(-\alpha)^{\uparrow\#b}.$$

(ii) Consider π as a partition of $u^{(1)}, \ldots, u^{(n)}$ where $u^{(i)} := \{(i-1)j+1, \ldots, ij\}$ for $i = 1, \ldots, n$ and put $\pi' = \{\cup_{k\in b} \cup_{l\in u^{(k)}} \sigma(l) : b \in \pi\}$, a j-even partition of $[nj]$ which has probability

$$\frac{n!}{(nj)!} \frac{(\theta/\alpha)^{\uparrow\#\pi}}{\theta^{\uparrow n}} \prod_{b\in\pi} -(-\alpha)^{\uparrow\#b} \frac{(j\#b)!}{\#b!}.$$

The partition π' determines an integer partition $(m_1, \ldots, m_{nj})$ of nj for which $m_i = 0$ for $i \notin \{j, 2j, \ldots, nj\}$. To each such integer partition there are $\frac{(nj)!}{\prod_{i=1}^{n}((ij)!)^{m_{ij}} m_{ij}!}$ associated even partitions with block sizes given by (m_i). Hence, we have

$$\mathbb{P}_{\alpha,\theta}(m_1, \ldots, m_{nj}) = \frac{n!(\theta/\alpha)^{\uparrow m_\bullet}}{\theta^{\uparrow n}} \prod_{i=1}^{n} \frac{[-(-\alpha)^{\uparrow i}]^{m_{ij}}}{(i!)^{m_{ij}} m_{ij}!}. \tag{8}$$

Note that (8) coincides with (4).

3.1 A Chinese restaurant construction for even partitions

For $n, j \in \mathbb{N}$, construct $\pi \in \mathcal{P}_{[nj]:j}$ according to the following random seating rule. Let (α, θ) be subject to the constraints of Pitman's two-parameter model.

(1) Units arrive in groups of size j;

(2) the first j units, $u_1, \ldots, u_j$, are seated at the same table;

(3) after nj arrivals are seated according to partition π, the next j units, $u_{nj+1}, \ldots, u_{(n+1)j}$ are seated as follows:

a. each unit u_{nj+i} for $i \geq 2$, chooses a unit $u^{(i)}$ uniformly among $u_1, u_2, \ldots, u_{nj+i-1}$ and switches positions with $u^{(i)}$;

b. $(u_{nj+1}, u^{(2)}, \ldots, u^{(j)})$ is treated as a single unit and chooses a table, $b \in \pi$, according to the two parameter CRP with parameter (nj, α, θ);

c. $(u_{nj+1}, u^{(2)}, \dots, u^{(j)})$ is seated at b.

As in section 2.2, we write $\pi \mapsto \pi'$ to indicate that $\pi' \in \mathcal{P}_{[(n+1)j]:j}$ is accessible from $\pi \in \mathcal{P}_{[nj]:j}$ according to the above seating rule.

Proposition 3.1. *The probability distribution on $\mathcal{P}_{[nj]:j}$ that results from the above construction is*

$$p_n^j(\pi;\alpha,\theta) = \frac{j^{\#\sigma-1}\Gamma(n)}{\Gamma(nj)} \frac{(\theta/\alpha)^{\uparrow\#\pi}}{(\theta/j)^{\uparrow n}} \prod_{b\in\pi} -(-\alpha/j)^{\uparrow(\#b/j)} \frac{\Gamma(\#b)}{\Gamma(\#b/j)}. \tag{9}$$

Proof. We show (9) by induction on n. Fix $j \in \mathbb{N}$ and (α, θ) in the parameter space of the two-parameter model, (3). Then (9) holds for $n = 1$ since $p_1^j(\pi;\alpha,\theta) = \mathbb{I}_{\{\#\pi=1\}}$, as it must according to step (1) of the above seating plan.

Now, assume that (9) holds for $n \in \mathbb{N}$ and consider $\pi^* \in \mathcal{P}_{[(n+1)j]:j}$. Let $A_{\pi^*} := \{\pi \in \mathcal{P}_{[nj]:j} : \pi \mapsto \pi^*\}$. By construction, the block sizes of each $\pi \in A_\sigma$ and π^* are identical except for the block to which the new group, $(u_{nj+1}, u^{(2)}, \dots, u^{(j)})$, is inserted in π during step 3c. to obtain π^*. Let this block be written as $b \in \pi$ and $b^* \in \pi^*$ so that $\#b = \#b^* - j$.

Let π' denote the restriction of π^* to $\{nj+2, \dots, (n+1)j\}$, regarded as a labeled partition with each $b \in \pi'$ labeled by $l(b)$ where

$$l(b) = \begin{cases} nj+1, & b \subset b^* \in \pi^* \\ \min\{i \in b\}, & b \subset b' \in \pi^*,\ b' \neq b^*. \end{cases}$$

In other words, each block, other than that corresponding to b^*, is labeled by the least element of the block to which it corresponds in π^*; and the block of π' which corresponds to b^* is labeled $nj+1$. This is a well-defined labeling since the least element of any block other than b^* in π^* must also be an element of the corresponding block in $\pi \in A_\sigma$, and $nj+1 \in b^*$ by definition.

According to the random seating plan, π' is obtained by displacement of the elements of π. Let $\varphi_{\pi,\pi^*} : [(n+1)j]\backslash[nj+1] \to [(n+1)j]$ be the operation corresponding to step 3a. of the random seating plan, i.e. for each $i \geq 2$, $\varphi_{\pi,\pi^*}(u_{nj+i})$ is the element in π which u_{nj+i} displaces in step 3a. of the random seating process. Let $\varphi_{\pi,\pi^*}(\pi')$ correspond to the labeled partition in which each element i of π' is replaced with $\varphi_{\pi,\pi^*}(u_i)$. There are $\binom{\#b+j-1}{j-1}$ such choices of the elements of $\varphi_{\pi,\pi^*}(\pi')$ and $\frac{(j-1)!}{\prod_{b\in\pi'}(\#b)!}$ ways to arrange them into a labeled partition with block sizes corresponding to the block sizes of π'. Multiplication of these factors gives $\frac{(\#b+j-1)\cdots(\#b+1)}{\prod_{b\in\pi'}(\#b)!}$ partitions in A_{π^*}. Furthermore, for each $\pi \in A_{\pi^*}$, the assignments under φ_{π,π^*} of the elements in each block of size r can be arranged in $r!$ possible ways. Hence, there are $(\#b+j-1)\cdots(\#b+1)$ possible ways for π^* to be obtained from a partition in $\mathcal{P}_{[nj]:j}$ through random seating.

Each random displacement in step 3a. has probability $\frac{1}{(nj+1)\cdots(nj+j-1)}$ and the random table assignment in step 3b. follows $\mathrm{CRP}(nj,\alpha,\theta)$ which assigns probability $\frac{\#b-\alpha}{nj+\theta}$ to b if $b \neq \emptyset$ and $\frac{\theta+\alpha\#\pi}{nj+\theta}$ if $b = \emptyset$.

By induction, (9) holds for all $n \geq 1$. $\square$

For $\alpha = -\kappa < 0$ and $\theta = m\kappa$ for some $m \in \mathbb{N}$, (9) is a distribution on $\mathcal{P}^{(m)}_{[nj]:j}$, j-even partitions with at most m blocks, which has limit

$$p^j_n(\pi;\lambda) = \frac{\Gamma(n)}{\Gamma(nj)} \frac{\lambda^{\#\pi} \prod_{b\in\pi} \Gamma(\#b)}{(\lambda/j)^{\uparrow n}}, \tag{10}$$

as $\kappa \to 0$ and $\theta \to \lambda$, the Ewens distribution with parameter λ/j restricted to j-even partitions.

Note that the distribution in (9) is the distribution of an $(\alpha/j, \theta/j)$-partition conditioned to be even of order j.

From (9), we obtain the combinatorial identity

$$\frac{1}{(nk)!} \sum_{\pi\in\mathcal{P}_{[nk]:k}} \alpha^{\#\pi}\Gamma(\pi) = \left(\frac{\alpha}{k}\right)^{\uparrow n}/n!, \tag{11}$$

where $\Gamma(\pi) := \prod_{b\in\pi} \Gamma(\#b)$.

3.2 Projective system associated with even partitions

Consider the two-step procedure at the beginning of section 3. The distribution of (π, σ) obtained in this way is

$$p_n(\pi, \sigma; \alpha, \theta) = \frac{(\theta/\alpha)^{\uparrow\#\pi}}{(nj)!\theta^{\uparrow n}} \prod_{b\in\pi} -(-\alpha)^{\uparrow\#b}. \tag{12}$$

We have shown that the even partition determined by (π, σ) where π is regarded as a partition of groupings $u^{(i)} : \{(i-1)j+1, \dots, ij\}$, $i = 1, \dots, n$, is distributed as (9), but this family is not defined on a projective system.

Each pair (π, σ) is an element of $\mathcal{P}_{[n]} \times \mathcal{S}_{nj}$, which is a product of projective systems, and hence is a projective system $\left(\mathcal{P}_{[n]} \times \mathcal{S}_{nj}, n \geq 1\right)$ with projection maps $\bar{D}_{m,n} := (D_{m,n}, \psi_{mj,nj})$, $m \leq n$, that act componentwise on $\mathcal{P}_{[n]} \times \mathcal{S}_{nj}$.

It is clear that the finite-dimensional distributions in (12) are invariant under the natural group action of $\mathcal{S}_n \times \mathcal{S}_{nj}$. Consistency is also straightforward and is now shown.

Proposition 3.2. *The finite-dimensional distributions in* (12) *on* $\mathcal{P}_{[n]} \times \mathcal{S}_{nj}$ *are consistent under* $\bar{D}_{m,n}$.

Proof. The form of (12) is the product of independent random objects $\pi \sim (\alpha, \theta)$ which is a consistent family on $(\mathcal{P}_{[n]}, n \geq 1)$ and a uniform permutation $\sigma \in \mathcal{S}_{nj}$, also consistent under $\psi_{(n-1)j,nj}$. Hence, the product distribution is trivially consistent with respect to $\bar{D}_{n-1,n}$. ∎

Kolmogorov's theorem immediately implies the existence of an infinitely exchangeable measure.

Theorem 3.3. *There exists an infinitely exchangeable measure* p^j *on* $\mathcal{P} \times \mathcal{S}$ *such that for* $p^j_n(\cdot; \alpha, \theta)$ *in* (12)

$$p^j_n((\pi, \sigma); \alpha, \theta) = p^j(\{(\pi^\infty, \sigma^\infty) : (\pi^\infty, \sigma^\infty)_{|[n]} = (\pi, \sigma)\})$$

for every $(\pi, \sigma) \in \mathcal{P}_{[n]} \times \mathcal{S}_{nj}$.

As in section 2.2, a sequence $(\pi_n, \sigma_n)_{n\geq 1}$ with each (π_n, σ_n) distributed as (12) determines a sequence $(\sigma_n \pi_n, n \geq 1)$ of even partitions with finite-dimensional distributions as in (9).

3.3 Even permutations

The distribution on pairs $(\pi, \sigma) \in \mathcal{P}_{[n]} \times \mathcal{S}_{nj}$ induces a distribution on $\mathcal{S}_n \times \mathcal{S}_{nj}$ in a straightforward way by choosing $\sigma_0 \in \mathcal{S}_n$ uniformly among the permutations of $[n]$ which correspond to $\pi \in \mathcal{P}_{[n]}$. This yields a distribution

$$\mathbb{P}_{\alpha,\theta}(\sigma_0, \sigma) = \frac{(\theta/\alpha)^{\uparrow \#\sigma_0}}{(nj)!\theta^{\uparrow n}} \prod_{b\in\pi} \frac{-(-\alpha)^{\uparrow \#b}}{\Gamma(\#b)}, \tag{13}$$

on the projective system $\mathcal{S}_n \times \mathcal{S}_{nj}$ with projection maps $\bar{\psi}_{m,n} := (\psi_{m,n}, \psi_{mj,nj})$.

The distribution induced by (13) on even permutations can be obtained directly from (9) and is

$$p_n^j(\sigma; \alpha, \theta) = \frac{j^{\#\sigma-1}\Gamma(n)}{\Gamma(nj)} \frac{(\theta/\alpha)^{\uparrow \#\sigma}}{(\theta/j)^{\uparrow n}} \prod_{b\in\sigma} \frac{-(-\alpha/j)^{\uparrow (\#b/j)}}{\Gamma(\#b/j)}. \tag{14}$$

4 Discussion

Here we have shown a projection of the well known Ewens process on set partition of $\mathbb{N}$ onto the spaces of balanced and even partitions, which have potential applications to complete and balanced block designs in statistics [1]. Our treatment shows a CRP construction on a product of projective systems which is straightforward in characterizing an infinite process, and which can be projected to their associated spaces of balanced and even partitions, which are not projective systems. Results of the projected process are obtained by implementing the usual combinatorial identities in this realm, and a potentially new combinatorial connection between partitions, even partitions and the gamma function is shown in (11).

References

[1] Bailey, R.A. (2008). *Design of Comparative Experiments*, New York: Cambridge.

[2] Billingsley, P. (1995). *Probability and Measure*, 3rd edition, New York: John Wiley.

[3] Burke, C.J and Rosenblatt, M. (1958). A Markovian Function of a Markov Chain. *Annals Math. Stat.* **29**(3), 1112-1122.

[4] Crane, H. (2011). A consistent Markov partition process generated from the paintbox process. *J. Appl. Prob.* **48,** to appear.

[5] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87-112.

[6] Kingman, J.F.C. *Mathematics of Genetic Diversity*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

[7] McCullagh, P. (2010). Random permutations and partition models. *International Encyclopedia of Statistical Science*.

[8] Pitman, J. (2005). *Combinatorial Stochastic Processes*, New York: Springer.